DOCUMENT RESUME

ED 410 245 TM 026 984

AUTHOR	Falk, Beverly; Ort, Suzanna
TITLE	Sitting Down To Score: Teacher Learning through Assessment.
PUB DATE	Mar 97
NOTE	20p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Conferences; Cooperation; *Educational Assessment; Elementary Secondary Education; Learning; *Participation; Performance Based Assessment; *Scoring; *Standards; State Programs; Teacher Attitudes; Teacher Education; *Teachers; Testing Programs
IDENTIFIERS	Large Scale Assessment; New York

ABSTRACT

This paper contributes to the body of knowledge on new forms of assessment by examining the impact of teacher involvement in the scoring of standards-based performance assessments. Researchers studied 250 of 500 teachers who participated in a New York state project to pilot new state assessment prototypes. The National Center for Restructuring Education, Schools, and Teaching (NCREST) collaborated with the New York State Education Department and other partners to develop new prototypes for the state assessment system that is standards-based. This project, the New York State Goals 2000 New Assessments Project mobilized teams of New York teachers to pilot the developed assessments and to attend scoring conferences to evaluate student work in relation to state standards. Data about 250 scoring conference participants came from observations during scoring sessions, interviews with 30 participants, and an open-ended questionnaire administered to all participants. Data provide strong evidence for how teacher involvement in the use of large-scale standards performance assessments supports teachers' learning about standards, their discipline, their students, and their teaching. It also demonstrates how the process of scoring offers teachers a forum for collaboration and learning from each other. (Contains 1 table and 33 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 410 245

Sitting Down to Score:
Teacher Learning Through Assment

Beverly Falk
Suzanna Ort

TM026984

This paper is prepared for the Annual Meeting of the American Educational Research Association in Chicago, IL

BEST COPY AVAILABLE

Sitting Down to Score: Teacher Learning through Assessment

Paper presented at the Annual Meeting of the American Educational Research Association
Chicago, Illinois
March 27, 1997

Beverly Falk and Suzanna Ort
National Center for Restructuring Education, Schools, and Teaching (NCREST)
Teachers College, Columbia University

Many current educational reform initiatives present the occasion, as well as the need, for new forms of professional development. Whether these reforms focus on issues of standards, curriculum and pedagogy, equity among diverse student populations, the nature and uses of student assessment, the social organization of schooling, or the professionalization of teaching, their success depends on the existence of teachers who can employ more complex skills and understandings than have ever been needed before. In addition to requiring teachers to know about disciplinary content, instructional strategies and teaching methodologies, all of these reforms call on teachers to utilize knowledge of learning theory, human development, personal and organizational change.

Recent research and commentary is building a base of understandings about new forms for professional learning that develop the capacities of teachers in these areas (Darling-Hammond, Ancess, and Falk, 1995; Fosnot, 1989; Giroux, 1988; Lieberman, 1995; Lieberman and McLaughlin, 1992; Little, 1992, 1993; Meier, 1992). This body of knowledge points away from the traditional short-term teacher "training" model that transmits information and skills to passive recipients toward a more long-range, capacity-building approach to professional development that offers "meaningful intellectual, social, and emotional engagement with ideas, with materials, and with colleagues both in and outside of teaching" (Little, 1993, p. 10).

Such an approach to teacher learning is a recurring theme in studies of teachers and schools who have been engaged in new strategies for assessing student work (Archbald and Newman, 1988; Andrias, Kanevsky, Strieb, and Traugh, 1992; Darling-Hammond and Ancess, 1994; Darling-Hammond, Ancess, and Falk, 1995; Falk, 1994; Falk, 1995; Falk and Darling-Hammond, 1993; Falk, MacMurdy, and Darling-Hammond, 1995; Weister and Yancey, 1996). All of these studies indicate that involvement in assessment activities stimulate professional development for teachers as well as organizational development for schools as a whole. Some of these studies reveal that looking at and deliberating on student work - with other teachers, students, and sometimes their families - helps teachers to think more deeply about their teaching and their students while, at the same time, enhancing their understanding of learning in general and of the strengths, styles, and needs of their individual students. Others of these studies suggest that engagement with new assessment strategies helps teachers develop a curricular vision for their teaching, as well as a focus on how to connect learners to those goals.

Draft: Please do not cite, duplicate or circulate without permission of the authors

BEST COPY AVAILABLE

Studies of large-scale assessment reform initiatives have contributed additional information about the impact of looking at student learning in relation to standards and goals. The standards-based performance assessments that are increasingly being used across the country appear to help teachers better understand what their students need to know and do. In addition, they stimulate teachers to think about changes in instructional approaches that can support students' learning (Falk and Larson, 1994; Koretz, Stecher, Klein, and McCaffrey, 1994; Sheingold, Heller, and Paulukonis, 1995).

This paper contributes to the growing body of knowledge by examining the impact of teacher involvement in the scoring of standards-based, performance assessments. We studied 250 of 500 teachers who participated in a New York State project to pilot new state assessment prototypes. Analysis of the data collected reveals that teacher involvement in the use and scoring of these new assessment forms supports teachers' learning, strengthens their sense of professionalism, and facilitates efforts to enact change.

The New York State Goals 2000 New Assessments Project: New Forms of Assessment, New Ways to Learn

In response to the recent national press to realize higher standards for student achievement, New York State, in 1991, launched a variety of initiatives to improve student learning. The agenda for change included articulating rigorous standards of achievement in a variety of disciplines; developing challenging curricula based on these standards; building the capacities of teachers to use a range of strategies to help students achieve the standards; and designing and using new forms of assessment that better support and reflect what is being taught (New York State Curriculum and Assessment Council, 1993).

As part of this initiative, the National Center for Restructuring Education, Schools, and Teaching (NCREST) collaborated with the New York State Education Department and other partners on a project to develop prototypes for the redesign of the state student assessment system. The redesign was intended to move the State from a testing program focusing on summative evaluation of curricula using primarily multiple choice forms of testing, to a standards-based system of performance assessments to be used for accountability purposes in the service of ongoing teaching and learning. This project - the New York State Goals 2000 New Assessments Project - mobilized teams of New York State teachers to design and construct standards-based performance assessment prototypes, to pilot these assessments in classrooms across the State, and to attend state-wide and regional scoring conferences where student work was evaluated in relation to state standards.

The assessment prototypes designed for this project measure the use of knowledge and skills in real-world contexts and applications, require higher-level thinking and complex problem-solving, and provide multiple ways for students to demonstrate their knowledge, skills, and understandings about many dimensions and kinds of learning. Much like the assessments that are used predominantly in other countries, these assessment prototypes consist mainly of extended essays and/or student-constructed responses that call on students to analyze, investigate, experiment, and present their findings in written, oral and/or graphic ways.

Draft: Please do not cite, duplicate or circulate without permission of the authors

BEST COPY AVAILABLE

Evaluation of student responses to the tasks is done by teachers using the clearly articulated criteria provided in each assessment's accompanying scoring guides or rubrics.

During the 1995-96 year of the Goals 2000 New Assessments Project, nine assessments in four disciplines were administered to approximately 12,000 elementary, middle, and high school students by 500 teachers representing over 100 districts across New York State. 250 of these participating teachers attended the project's end-of-year scoring conference. The purpose of the conference was to score student work in relation to standards in order to generate data for three purposes: 1) to analyze the technical merits of the assessments - their overall validity and ability to measure student progress reliably; 2) to analyze teacher and student reactions to the new assessment prototypes; 3) to examine the professional development potential of a standards-based performance assessment system in use (National Center for Restructuring Education, Schools, and Teaching, 1997).

This study reports on findings related to the third purpose of the scoring conference. [Technical analyses of the assessments as well as analyses of teachers' and students' responses to the new assessments are reported elsewhere (National Center for Restructuring Education, Schools, and Teaching, 1996)]. Data for the study was obtained from three sources: from observations of groups of teachers during scoring sessions; from interviews with 30 teachers attending the scoring conference (it should be noted that while these teachers were representative of the geographic and ethnic diversity of the state, they may not be fully representative of the views of all teachers participating in the project because they volunteered to participate in the interviews); and from an open-ended questionnaire administered to all conference participants about the effects of examining student work in relation to standards.

Scoring Student Work: A Learning Experience

Building on New York State's long tradition of involving teachers in the development and evaluation of state examinations, the standards-based performance assessment prototypes developed through the Goals 2000 New Assessments Project relied on teachers to score exams. This was done through the use of scoring guides or rubrics that articulate the essential criteria of the standards embodied in each assessment task and that describe indicators of performance for each criteria at four or five levels of proficiency. (See attached sample rubric.)

The scoring process took place in discipline-based, grade-level groups. First, teachers reviewed each assessment task's rubrics to discuss the essential criteria embodied in each task and the rubrics' descriptions of student performance at differing levels of proficiency. Then the teachers read sample responses to exam questions with the goal of assigning each a score. In groups of six to eight teachers, they compared the evidence in each response to the rubric's indicators of different performance levels. They discussed the work, detail by detail, until they arrived at a consensus for a score. Inevitable differences in opinions and perspectives were mediated by the scoring protocol's reliance on the rubric's clearly articulated criteria for performance and the insistence that teachers always justify their evaluation using evidence from the student work.

Draft: Please do not cite, duplicate or circulate without permission of the authors

BEST COPY AVAILABLE

After several hours of this practice orientation, teachers began to consistently agree on the scores they assigned to student responses. This recurring achievement of consensus signalled the completion of the orientation. Teachers then began the independent scoring and blind rescoring of tasks that would be used to assign the "official" score.

Because of this attention to evidence and the careful preparation of the teachers involved, the scoring process produced quite reliable student scores. Interrater agreements (between two different scorers evaluating student work independently) ranged from 80 to 94 percent, considered in large-scale measurement to be a quite strong evidence of reliability. In addition to producing reliable scores, the scoring process provided teachers with a structure to look closely at student work in collaboration with their colleagues. It also offered them an opportunity to inquire and learn about the New York State standards, their discipline, their students, and their teaching.

Clarifying goals and expectations for teaching and learning

An overwhelming majority of the teachers responding to a questionnaire about the New Assessments Project reported that the most valuable aspect of their participation in the project was the conversations they had with other colleagues about standards, assessments and student work. These discussions about standards provided them with coherence and continuity for their teaching as well as a guide to what is valued and valuable in the broader New York State community.

> The discussion about the scoring rubrics was extremely useful. It was a great way of sharing ideas on what people think is important. (Middle school mathematics teacher)

Interviews and observations of teachers during the scoring sessions supported this finding. As teachers at the scoring conference reviewed the standards together with other colleagues, they were both challenged and supported to learn. They learned about state expectations for their students, were presented with opportunities to consider how these expectations might affect their teaching, and, in the process, clarified how their own views differed or agreed with the state's. Since the standards are, in a sense, political documents, reflecting a consensus opinion of diverse constituents, working with the standards challenged teachers to arrive at their own balance between individual and group perspectives.

Deepening teacher knowledge of the disciplines

Scoring student responses to the assessment tasks offered teachers a way to see how the big ideas of the standards actually play out in real work. Just as assessment tasks embody standards, rubrics frame each task's criteria around important qualities of a discipline. The rubrics for the English Language Arts assessments, for example, present five essential qualities or dimensions of the discipline. (See attached sample.) These include understanding and analysis, idea development, organization, language use, and conventions (mechanics such as spelling, punctuation, grammar, etc.). Four different levels of proficiency are described for

Draft: Please do not cite, duplicate or circulate without permission of the authors

BEST COPY AVAILABLE

each one of these dimensions. As teachers throughout the scoring process referred to these dimensions and to the descriptions of how the dimensions were reflected in student work at different levels of proficiency, they developed shared understandings and a common language about the essentials of their discipline.

> Because of the scoring rubric, we, the scorers, were actually able to see and understand the qualities that were used to assess our students' work. (English Language Arts high school teacher)

> I found the scoring rubric to be extremely helpful and to be filled with implications of what is involved in the discipline. (English Language Arts high school teacher)

The rubric's frame for assessing student work guided teachers to use evidence, rather than personal response, as the basis for making evaluations.

> I moved away from thinking about work in an A, B, C, or D way, to thinking about the criteria for performance and the evidence that would justify my evaluation. (English Language Arts elementary teacher)

> The score wasn't an arbitrary decision. It was based on assessment of a "real" quality. (English Language Arts high school teacher)

> Using a well written rubric makes assessment as objective as possible. (English Language Arts elementary teacher)

Learning about students and their work

Looking at student work in relation to standards helped many teachers gain understandings about the strategies and approaches students bring to their learning. Of course, the ability to do this is contingent upon having worthy assessment tasks - tasks designed to be as contextualized as possible, to ask students to show and explain their work, and to allow a wide range of students to be able to demonstrate their abilities. When the tasks possess these characteristics, the process of scoring student responses offers a window into *what* their students know and can do as well as *how* their students actually do it.

Many teachers in the New Assessments Project credited the "authentic" aspects of the assessments with helping them gain such insights.

> When students are expected to explain or support their answers, you begin to learn more about what they understand. (Middle school mathematics teacher)

> [Scoring student work] gave me tremendous insights into the thinking of a child. (Elementary mathematics teacher)

Draft: Please do not cite, duplicate or circulate without permission of the authors

BEST COPY AVAILABLE

> I began to look more in depth for what the children were able to do. (English Language Arts high school teacher)

This deeper look at children's thinking, revealed through examining responses to the tasks, helped many teachers appreciate the variety of ways students solve problems and express their ideas.

> I learned that kids see things differently and perceive things differently. (High school mathematics teacher)

> I saw that students can really see the questions in a variety of ways. (Middle school teacher who administered a math/science/technology task)

Many teachers gained a deeper respect for formerly unnoticed students' strengths.

> The clear description of the expectations for the tasks helped me to see that some work is more acceptable than an initial reaction might give. It helped me in seeing what students *could* do, rather than only focusing on what they *couldn't* do. (English Language Arts high school teacher)

Recognizing differences in students' responses helped some teachers broaden their views of what constitutes "good work."

> I am more open to the variety of student work. (English Language Arts elementary teacher)

> My range of "good" expanded. In reviewing work I was able to see that students use many different strategies to problem-solve. (Elementary mathematics teacher)

> [Looking at student responses to the assessment tasks] reinforced the idea that good work can look very different and can take on many forms. (Elementary teacher who administered a math/science/technology task)

Teachers' visions of possibility for "good work" was expanded in yet another way throught the process of scoring assessments in collaboration with other colleagues. Because teachers were looking at student work representing a variety of geographic areas, types of locales, racial, ethnic, socioeconomic, and linguistic backgrounds from all across New York State, they were able to see what student work looked like beyond the confines of their classrooms, schools, and districts. This larger context for thinking about and evaluating work, broadened perspectives on the possible range of student performance.

Draft: Please do not cite, duplicate or circulate without permission of the authors

BEST COPY AVAILABLE

> I learned so much from seeing the work of students other than my own. (English Language Arts high school teacher)

> The most valuable part of the scoring session was learning about the differences and similarities of the populations of the disparate school districts. (English Language Arts high school teacher)

Additional benefits of scoring noted in teachers' responses to survey and interview questions were attributed to the nature of the new tasks. Most commonly cited as enhancing teachers' insights about students was the capacity of the new assessments to provide students with a variety of ways to demonstrate their knowledge and to explain their thinking.

> There is a chance for every kid to contribute something. Even if they only comprehended on the most minimal level, there's still a place for them to put down what they understood. (English Language Arts elementary teacher)

Especially noted was the new assessments' ability to provide all students with opportunities to reach toward and demonstrate higher standards.

> This test challenges all levels of students to perform at high standards. (Elementary mathematics teacher)

> I learned that low-achieving students can experience success as well as express an element of critical thought that I never believed possible. (English Language Arts high school teacher)

> I learned that all students can achieve. When the expectations are very clear, it seems students will meet them. (English Language Arts high school teacher)

Because the new assessments provide a broader range of students with opportunities to demonstrate their proficiencies, many teachers saw the new assessments as being fairer than other tests.

> I have changed my attitude. This is a fairer way of assessing. (English Language Arts elementary teacher)

Discussions of fairness led some teachers to consider broader issues of equity. As they discussed how to prepare students for these new exams, awareness heightened about the need for all students to have equitable opportunities to learn.

> [This experience] has made me see that we need to be striving for all students to achieve this good work. (Middle school mathematics teacher)

BEST COPY AVAILABLE

Draft: Please do not cite, duplicate or circulate without permission of the [illegible]

> This allows many entry points for kids at all levels. Hopefully, this will encourage middle and high schools to begin detracking and allowing all children access to the same curriculum and standards. (English Language Arts elementary teacher)

> This is a good start and will help raise standards and expectations for our students (Elementary mathematics teacher)

Developing insights to support teaching

Many teachers who took part in the scoring conference reported that the pre-scoring conversations and the actual scoring provided them with insights about how to improve their curriculum and teaching. Many felt that their understandings of what students should know and be able to do were strengthened in ways that would be useful for developing more cohesive programs of instruction.

> This project has made me more aware of expectations for good teaching and learning. It is a catalyst for planning. (English Language Arts elementary teacher)

> Piloting these assessments has made me focus on the real learning in the classroom - students' ability to think, analyze, and write. (English Language Arts high school teacher)

Other teachers pointed to scoring's usefulness in helping them focus on needed changes in their practice.

> I think I'm going to be teaching a little differently now: Group work activities; making sure students understand the concepts of what they're doing; not looking for just the wrong or right answer. That's very important and so is how students are arriving at their answer. (Elementary mathematics teacher)

> [Using this assessment] has motivated me to offer "more" in my classroom - more time for the students to express themselves; more opportunities to investigate (less follow me); more why and why not (less pushing to the "right" answer). (Middle school mathematics teacher)

> I will try to take into consideration the different degrees of ability the students have. I will try to be more supportive. (Middle school mathematics teacher)

> This assessment helped me to reflect on what I need to do. It is an enormous stimulus to far better practice. (English Language Arts elementary teacher)

Draft: Please do not cite, duplicate or circulate without permission of the authors

BEST COPY AVAILABLE

Some teachers credited the process of looking at student work in relation to standards as helping them become more aware of specific content areas in need of more focus.

> Working on this pilot has encouraged me to take a closer look at the way I teach math. (Elementary mathematics teacher)

> I need to focus instruction more on problem-solving content. (Middle school mathematics teacher)

> It gave me a reason to rethink how my units for science will be devised and implemented and I know that more problem-solving activities/assessments will be part of the process. (Elementary teacher who administered a math/science/technology task)

> It confirms and supports the way I teach and makes me realize that I need to keep asking kids for details - getting them to find the specific evidence from their reading (English Language Arts elementary teacher)

As a result of exposure to the approach and format of these prototype assessments, quite a few teachers decided to institute changes in their classroom-based assessment practices:

> I plan to give kids rubrics - maybe written - with what makes "quality" work. (English Language Arts elementary teacher)

> I will provide more opportunities for revision, self analysis, and evaluation. (English Language Arts elementary teacher)

> I am sure to make any open-ended questions I ask clear enough to get the information I want to get from the students. I will also make grading criteria very clear, very related to the question, and made available to students ahead of time. (Middle school teacher who administered a math/science/ technology task)

> I will do more testing requiring justifications and explanations - help students to become more comfortable explaining their understandings. (High school mathematics teacher)

> I plan to set up diagnostic conferences to help my students focus on their strengths and weaknesses. (English Language Arts elementary teacher)

> The first thing I will do next year is to give this assessment to my class. I plan to use this test as a guide for my language arts program. (English Language Arts elementary teacher)

BEST COPY AVAILABLE

Draft: Please do not cite, duplicate or circulate without permission of the authors

Strengthening teachers' sense of professionalism

The focused conversations that took place at the scoring conference before teachers individually sat down to score, offered opportunities for teachers to collaborate, to learn from other teachers, and to validate their knowledge as competent professionals.

> Meeting with dedicated, concerned teachers was most valuable to me. I learned from their positive attitudes, from discussing fears and concerns about my students, and the future of my discipline in this state. (English Language Arts high school teacher)

> It was very helpful to work with teachers from throughout the state, to hear their thoughts and discuss how to raise standards for students. (High school social studies teacher)

Evaluating student work in collaboration with colleagues provided teachers with opportunities to learn - to discuss, debate, and share ideas about what is important and how it is demonstrated. Many praised it as effective professional development.

> The scoring session provided valuable professional dialogue. It was a great way to do teacher inservice. (High school social studies teacher)

The process of scoring thus provides teachers with both a structure and a process to carry on the critical dialogue needed to enhance professional growth. It is a mechanism for keeping teacher inquiry ongoing and alive, an essential ingredient for the success of any effort to improve schooling (Berman and McLaughlin, 1977, 78; Cohen and Hill, 1997; National Commission on Teaching and America's Future, 1996).

Facilitating change

> There is no question in my mind that implementing a test similar to this will result in changes institutionally and in the classroom. More than anything else I've seen or heard in recent years, this test will bring about important and needed instructional changes. (English Language Arts elementary teacher)

Teacher involvement in the scoring of standards-based performance assessment promises to be a powerful catalyst for meaningful change. Because it brings teachers together to focus on and engage in meaningful discussions about students and their work, it strikes at the core of teaching and learning (Elmore, 1996). Scoring provides teachers with the time and a way to build their capacities to knowledgeably engage in new practices.

> I don't think you can underestimate the need that folks have for getting together and having quality time to reflect mutually on these changes. (High school social studies teacher)

BEST COPY AVAILABLE

Draft: Please do not cite, duplicate or circulate without permission of the authors

Only through such extended and meaningful experience with new practices can teachers and other stakeholders come to understand and own new initiatives. This is critical for realizing positives effects from new practices for students.

> Standards documents, even elegant ones with benchmarks and commentary, can affect achievement only if the standards come to be held as personal goals by teachers and students...That will happen only if a concerted effort is made to engage teachers and students in a massive and continuing conversation about what students should learn, what kinds of work they should do, and how well they should be expected to do it (Resnick, 1995, p. 113)

Involving teachers in the scoring of large scale assessment encourages them to be active participants in shaping the direction of school improvement efforts. It acknowledges their critical role in educational reform and puts them in their rightful place at the center of the change process.

> This whole process allows teachers to be part of a process of teaching and learning. Anything that can be explained through a process makes it easier for the learner. It allows me to be an active participant with a voice to question and make decisions. (English Language Arts high school teacher)

Concluding Thoughts: Benefits/Worries/Unanswered Questions

The data presented in this study provides strong evidence for how teacher involvement in the use of large scale standards-based performance assessments supports teachers' learning about standards, their discipline, their students, and their teaching. It also demonstrates how the process of scoring offers teachers a forum to collaborate and learn from each other.

While opportunities for learning in the arena of standards-based performance assessment are plentiful, along with these opportunities, come a variety of worries as well as some unanswered questions. (The Chinese ideograph for the word "opportunity" also signifies the word "danger." Standards-based performance assessment case is an excellent case in point.)

One worry is that the benefits of using standards-based performance assessments are dependent on the nature of the standards and the assessments. When standards emphasize challenging and dynamic aspects of learning and articulate core ideas and critical skills in and across disciplines without being overly prescriptive, they offer the opportunity for educational communities to clarify expectations as well as to develop shared meanings and language about these understandings. Useful standards can provide a helpful guide for teaching, can promote richer learning contexts, and thus encourage better student learning. The converse, however, is also true. If standards and their accompanying assessments repeat the mistakes of the 1970's competency-based education movement by specifying hundreds of discrete objectives that reduce subject matter to tiny subskills

Draft: Please do not cite, duplicate or circulate without permission of the authors

BEST COPY AVAILABLE

and facts, they may be too reductionistic to support thoughtful teaching. So - Caution Number One: Standards and their accompanying assessments need to be a rich and clear guide to practice that supports meaningful learning. In other words, standards need to be of a high standard.

The clarity of purpose made visible by standards-based assessments can be a powerful support to enhanced professional knowledge and a strengthened sense of professionalism. Yet, while clarified images of excellence can be beneficial, they can, at the same time, constrain possibilities for excellence (i.e., what if someone were to do something unusual and creative that was outside the bounds of the assessment criteria's description of excellence?) We must remain cognizant of this possibility and try to ensure, when developing standards and assessments, that there is room for the development of public consensus about quality work as well as the nurturance of individuality. So - Caution Number Two: Seek to develop standards and assessments that balance the tension between conformity, community, and individuality.

Issues of social justice and equity are also raised by the move toward standards-based assessments. Standards and standards-based assessments, with their clear criteria and public performance indicators, provide all teachers and students with explanations of and access to images of excellence. This open and public guide to teaching and learning offers a way to level the playing field a bit between students who have had vastly unequal opportunities, resources, and supports. Our accompanying worry however is that this equity benefit cannot be sufficiently realized unless adequate resources and supports are provided to build the capacities of *all* teachers and students to achieve the new more challenging standards. Special attention and additional resources need to be allocated to those students and groups of students who historically have had inequitable opportunities to learn. So - Caution Number Three: Equitable and adequate opportunities to learn must be the cornerstone of efforts to realize high standards.

A final caution has to do with how new standards and assessments are implemented and used. Standards-based performance assessments, regardless of their quality, might produce unintended consequences if they are too tightly coupled with high stakes. There is a danger that the learning value of these assessments can be seriously compromised if the assessments are used for sanctions, rewards, or other such high stakes purposes (Keene, 1997; Koretz, 1996; Linn, 1996). Studies of human motivation also caution us that extrinsic motivations, such as high stakes, can potentially harm intrinsic motivations. This may possibly have the effect of undermining original efforts aimed at supporting student learning (Kelleghan, Madaus, and Raczek, 1996; LeMahieu, 1996). So - Caution Number Four: To maximize the learning potential of assessments, minimize stakes as much as possible.

Now for the unanswered question(s): While this study provides evidence for how examining student work in relation to standards supports teachers' learning,

Draft: Please do not cite, duplicate or circulate without permission of the authors

BEST COPY AVAILABLE

what remains unclear is how these learnings actually translate to improved teaching. Can teachers actually incorporate their learnings from assessment into their teaching strategies? Will student performance actually improve as a result of what teachers have learned? Further studies are needed to explore how teachers use the understandings they gain through the scoring of standards-based performance assessments to strengthen teaching and learning in classrooms and to improve student achievement.

Implications for further work

Efforts to create and use standards-based performance assessment systems present major conceptual and implementation challenges pertaining to professional development. The conceptual challenge is to assert an intricate connection between teaching, learning, and assessment. The shift away from machine-scorable, multiple choice or fill-in-the-blank tests that have dominated this country throughout most of this century, toward standards-based, performance assessments that are evaluated by teachers in collaboration with their colleagues, presents an opportunity for the teaching profession to reclaim assessment from outside evaluators and experts. Performance assessments not only provide more direct and valid information about student progress than traditional assessments have ever offered, they yield information that is useful to teaching through a process that both enhances and validates teachers' knowledge. Performance assessments have the potential to be a powerful link between instruction, assessment, student learning, and professional development. In this coming decade, when it is predicted that up to 50 percent of the current teaching force will be replaced with new recruits, teacher involvement in assessment use provides a much needed opportunity for professional development

Realizing the benefits of large-scale standards-based assessments also presents an implementation challenge. The findings from this study shed some insight into how to address this operational issue. They suggest that teacher involvement in scoring can provide an efficient use of resources. Through one concentrated investment, assessment addresses two important needs: activities and expenditures used for assessment also yield professional development. In these times of ever-constraining resources, this is an attractive attribute. Teacher involvement in the scoring of assessment can address system accountability needs while offering a powerful way to affirm professional knowledge and to support teacher learning.

BEST COPY AVAILABLE

Draft: Please do not cite, duplicate, or circulate without permission of the authors

References

Andrias, J., Kanevsky, R.D., Strieb, L.Y., and Traugh, C. (1992). Exploring values and standards: Implications for assessment. New York: National Center for Restructuring Education, Schools, and Teaching.

Archbald, D.A., and Newman, F.M. (1988). Beyond standardized testing: Assessing authentic academic achievement in the secondary school. Reston, VA: National Association of Secondary School Principals.

Berman, P., & McLaughlin, M. (1977). Federal programs supporting educational change: Vol. VII. Factors affecting implementation and continuation. Santa Monica, CA: Rand Corporation.

Berman, P., & McLaughlin, M. (1978). Federal programs supporting educational change: Vol. VIII. Implementing and sustaining innovations. Santa Monica, CA: Rand Corporation.

Cohen, D. & Hill, H. (1997). Policy, practice and performance: The effects of policy on teaching and learning. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, Illinois, March, 1997.

Darling-Hammond, L., and Ancess, J. (1995). Graduation by portfolio at Central Park East Secondary School. New York: National Center for Restructuring Education, Schools, and Teaching.

Darling-Hammond, L., Ancess, J., and Falk, B. (1995). Authentic assessment in action. New York: Teachers College Press.

Elmore, R. (1996). Getting to scale with good educational practice. Harvard Educational Review, 66(1), 1-26.

Falk, B. (1996). Teaching the way children learn. In Milbrey McLaughlin and Ida Oberman (Eds.) Teacher learning. New York: Teacher College Press.

Falk, B. (1995). Authentic assessment as a catalyst for learning: An inquiry model of professional development. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, California.

Falk, B. (1994). The Bronx New School: Weaving assessment into the fabric of teaching and learning. New York: National Center for Restructuring Education, Schools, and Teaching.

Falk, B. and Larson, J. (1995). An invitation to invention: Top-down support for bottom-up reform of assessment in New York State. Paper presented at the

BEST COPY AVAILABLE

Annual Meeting of the American Educational Research Association, San Francisco, California.

Falk, B., MacMurdy, S., & Darling-Hammond, L. (1995). Taking a different look: How *The Primary Language Record* supports teaching for diverse learners. New York: National Center for Restructuring Education, Schools, and Teaching.

Fosnot, C. T. (1989). Enquiring teachers, enquiring learners. New York: Teachers College Press.

Giroux, H.A. (1988). Teachers as intellectuals: Toward a critical pedagogy of learning. Granby, Mass.: Bergin and Garvey.

Keene, A. (1997). Improving assessment so all children reach high standards: The classroom perspective. Presentation at "Examining the Multiple Purposes of District-Level Assessment," a Special Interest Workshop sponsored by the Region III Comprehensive Center and the Center for Excellence and Equity in Education, George Washington University, February 20, 1997, Alexandria, VA.

Koretz, D. (1996). Summary of findings: Technical Advisory Group for the New York State New Assessments Project. New York: The National Center for Restructuring Education, Schools, and Teaching.

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. Educational measurement: Issues and practice, 13, 5-16.

LeMahieu, P. (1996). Summary of findings: Technical Advisory Group for the New York State New Assessments Project. New York: The National Center for Restructuring Education, Schools, and Teaching.

Lieberman, A. (1995). Practices that support teacher development: Transforming conceptions of professional learning. Phi Delta Kappan, 76(8), 591-596.

Lieberman A. and McLaughlin, M.A. (1992). Networks for educational change: Powerful and problematic. Phi Delta Kappan, 73(9), 673-677.

Linn, R. (1996). Summary of findings: Technical Advisory Group for the New York State New Assessments Project. New York: The National Center for Restructuring Education, Schools, and Teaching.

Little, J.W. (1993). Teachers' professional development in a climate of educational reform. New York: National Center for Restructuring Education, Schools, and Teaching.

Draft: Please do not cite, duplicate or circulate without permission of the authors

BEST COPY AVAILABLE

Little, J.W. (1992). Teacher development and educational policy. In M. Fullan & A. Hargreaves (Eds.), Teacher development and educational change, pp. 170-193. London: Falmer Press.

Madaus, G. (1996). The use of external examinations to improve student motivation. Washington, D.C.: American Educational Research Association.

Meier, D. (1992). Reinventing teaching. Teachers College Record, 93(4), 594-609.

National Center for Restructuring Education, Schools, and Teaching. (1997). The New York State Goals 2000 New Assessments Project: Technical Report. New York: National Center for Restructuring Education, Schools, and Teaching.

National Commission on Teaching and America's Future. (1996). What Matters Most: Teaching for America's Future. New York: National Center for Restructuring Education, Schools, and Teaching.

New York State Curriculum and Assessment Council. (1993). A learning-centered curriculum for New York State. Albany: New York State Education Department.

Resnick, L. (1995). Standards for education. In D. Ravitch (Ed.), Debating the future of American standards. Washington, D.C.: The Brookings Institution.

Sheingold, K., Heller, J.I., & Paulukonis, S.T. (1995). Actively seeking evidence: Teacher change through assessment development. (Center for Performance Assessment Report MS #94-04). Princeton, NJ: Educational Testing Service.

Thomas, W., Storms, B., Sheingold, K, Heller, J., Paulukonis, S., Nunez, A., & Wing, J. (1995). California Learning Assessment System portfolio assessment research and development project: Final report. Princeton, N.J.: Center for Performance Assessment, Educational Testing Service.

Weister, I., & Yancey, K. (In press). Teachers and students: Reclaiming assessment. Portfolios for Learning and Beyond. Urbana, IL: National Council of Teachers of English.

BEST COPY AVAILABLE

Student I.D. Number:_____ Reading Comprehension Score: _____

Scorer #1 I.D. Number:_____ Writing Effectiveness Score: _____

Scoring Rubric for New York State Elementary English Language Arts Assessment Spring 1996

Task 3: Reading and Writing for Information and Understanding and for Critical Analysis and Evaluation in an Informational Text

Dimensions	Descriptors 4 Accomplished	3 Proficient	2 Developing	1 Beginning
Reading Comprehension				
Understanding	Includes a broad range of critical information that explains how and why each animal in each category disguises itself. Elaborates on and thoroughly explains information from the text.	Includes critical information that explains how and/or why each animal in each category disguises itself with some elaboration and explanation.	Presents some information with some explanation. May contain limited factual errors and/or misinterpretations.	Presents information with limited or no explanation. May contain factual errors and/or misinterpretations. May be missing critical information.
Analysis	Goes beyond the factual information presented in the text to interpret and analyze.	Some evidence of interpretation and analysis.		
Writing Effectiveness				
Idea Development	Develops and elaborates ideas clearly and fully using many supportive and relevant details from the text. Draws meaningful connections between ideas.	Develops ideas clearly with some supporting details from the text.	Ideas are stated simply. Few supporting details from the text are referenced. May wander from the topic or task.	Develops ideas in fragmentary manner. Does not use supporting details from the text and/or includes random information and personal details unrelated to the topic. May be off-topic or task completely.
Organization	Writing has a distinctive organizing shape and structure. Ideas are presented in a clear order and logical sequence with transitions made through the use of paragraphing, introductions, and/or conclusions.	Ideas are presented in a clear order and logical sequence. Writing is easy to understand and follow.	Ideas are presented in some sequence but overall coherence is tentative. Response may require several rereadings to understand what is written.	Ideas are presented with little or no organizational pattern. Difficult to follow.
Language Use	Uses lively and descriptive language that goes beyond what was stated in the text. Engages the reader, has a sense of audience and a distinguishing sense of voice. Varied sentence length and structure.	Uses language and vocabulary from the text. Writes to an appropriate audience. Some sense of voice and sentence variety.	Word choices are simple, as are sentences, with little attention paid to audience.	Vague or unclear language (hard to understand).
Conventions	Writing is generally free of errors in spelling, punctuation, and grammar.	Limited errors in spelling, punctuation, and grammar but the meaning of the writing is communicated.	Numerous errors in spelling, punctuation, and grammar but the meaning of the writing is distinguishable.	Mechanics interfere with the reader's ability to understand the response.

BEST COPY AVAILABLE